Temporal Models for Robot Classification of Human Interruptibility

Siddhartha Banerjee
Georgia Institute of Technology
Atlanta, GA 30332, USA
siddhartha.banerjee
@gatech.edu

Sonia Chernova
Georgia Institute of Technology
Atlanta, GA 30332, USA
chernova@gatech.edu

CSCW'17, Feb 25-Mar 1, 2017, Portland, Oregon USA
ACM .

Abstract

Robots are increasingly deployed to human work environments where they will need to approach and interrupt collocated humans. Prior research has shown that mistimed interruptions can have a large negative impact on human task performance. Our aim is to equip a robot with the wherewithal to generate socially acceptable interruptions so that any negative effects of the robot interruptions are mitigated. This work makes three contributions to the research area. First, we introduce an ordinal scale that can be used to rate the interruptibility of a human. Second, we propose the use of Conditional Random Fields (CRFs) and their variants, Hidden CRFs, and Latent-Dynamic CRFs, for classifying interruptibility. Third, we introduce the use of object labels as a visual cue to the context of an interruption in order to improve interruptibility estimates. Our results show that Latent-Dynamic CRFs outperform all other models across all tested conditions, and that the inclusion of object labels as a cue to context improves interruptibility classification performance, yielding the best overall results.

Author Keywords

Human-Robot Interaction; Interruptibility; Conditional Random Fields

ACM Classification Keywords

I.2.m [Artificial Intelligence]: Miscellaneous

Introduction

Robots are increasingly deployed to human work environments in which they will need to approach collocated humans, whether to signal task completion, to report a problem, to request help, or to offer a service. These interactions will often serve as interruptions for the humans involved, who might already be engaged in other tasks. The goal of this project is to equip robots with the wherewithal to engage people but cause minimal disruption to the person's ongoing task.

Psychology research has shown that poorly timed interruptions can have negative impacts on human task performance [4], which in turn has prompted research into a measure called *interruptibility*. The interruptibility of a person at any given point in time is defined as a measure of the receptiveness of that person to external disturbances (interruptions) at that moment [15]. Low interruptibility signifies the person's desire not to be disturbed, while high interruptibility signifies that the person could be amenable to an interruption. Humans are very adept at gauging the interruptibility of others from observation in the workplace [10] and we rely on that knowledge to equip a mobile robot to do the same.

Our work makes three contributions to this research area. First, we introduce an ordinal scale of interruptibility that can be used to rate the interruptibility of a person and to influence decisions on whether or not to interrupt. Second, we explore the use of Conditional Random Fields (CRFs) [6] and their variants, Hidden CRFs (HCRFs) [16], and Latent-Dynamic CRFs (LDCRFs) [9], for classifying interruptibility. Using a dataset of person observations collected by a mobile robot, we compare the performance of these models against HMMs and show that the LDCRF consistently outperforms all other models across all tested conditions. Third, motivated by work on interruptibility in other areas of computing [17], we introduce the use of object labels as a visual cue to the *interruption context* in order to improve interruptibility estimates. Our results show that inclusion of object labels as a cue to context improves interruptibility classification performance, yielding the best overall results.

Related Work

In addition to research in the related field of Ubiquitous Computing (Ubicomp) [17], robotics research has looked at the problem of interruptibility through the work of Rosenthal et al. [11], Satake et al. [12], and Shi et al. [14]. Rosenthal et al. used accumulated knowledge on the occupancy schedule of people in offices to predict availability, under the assumption that the person was interruptibile if their door was open. In our work, we do not make such an assumption and rely on information about the person at the desired moment of interruption to judge interruptibility. Shi et al. and Satake et al. relied on extensive instrumentation of the environment to track people to train SVMs and create models for when to engage a human. However, we wish to predict interruptibility for a mobile robot in an unstructured environment and therefore we do not rely on information from external sensors.

There has been related research on detecting human engagement [8] and on estimating a human's awareness of a robot [2]: insights from this research forms the baseline for our work. Specifically, the work of Chiang et al. [2] and Mollaret et al. [8] models human intent-to-engage and human awareness using HMMs trained on audio features, body position features, and head gaze features. We leverage their approach for the classification of interruptibility.

Additional research has focused on determining the best

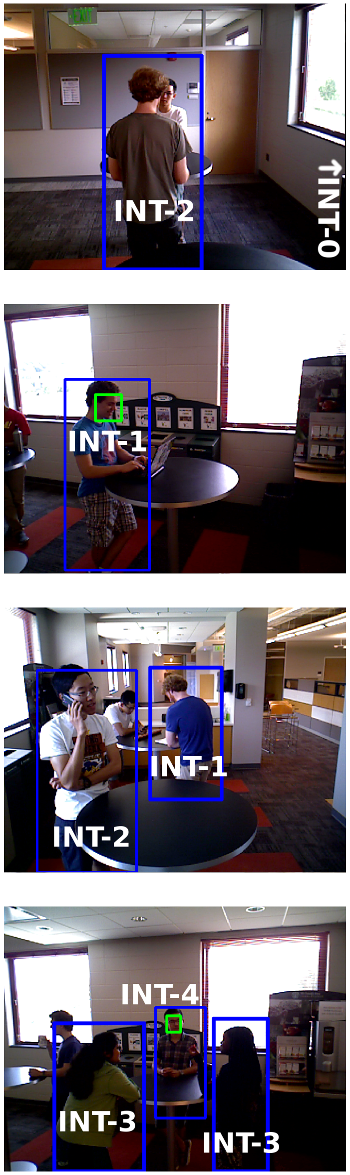


Figure 1: Example scenes from data collection in a shared kitchen area. The green bounding box denotes a face identified by the face recognition component and the interruptibility label of individuals within the blue bounding box is shown.

way to perform the interruption once a human is defined as interruptible. Saulnier et al. [13] have explored the most appropriate set of nonverbal behaviours for interruptions while Chiang et al. [2] have used Reinforcement Learning to personalize interruption behaviors. Our work focuses on the first part of the problem, which is to classify whether the person is interruptible at a point in time.

Finally, prior work highlights the importance of *interruption context* [10, 17]. Computationally, the interruption context broadly consists of features that describe the user (e.g., personality traits) [15], the task [1], the environment [3], the interruption [5], and the relationships between these [7]. In our work, we focus on garnering environment context: we hypothesize that the labels of objects that the person is interacting with can serve as valuable contextual cues to improving classification of interruptibility.

Methodology

In this section we introduce the ordinal scale of interruptibility, describe the types of features that we used, introduce the four temporal models for classification, and briefly describe the data collection and annotation effort.

Interruptibility Scale

We propose the following scale to classify interruptibility:

INT-4 **Highly Interruptible**. The person is not busy and they are aware of the robot's presence.

INT-3 **Interruptible**. The person is not busy, but they are unaware of the robot's presence.

INT-2 **Not Interruptible**. The person is busy, but the robot may interrupt if necessary.

INT-1 **Highly Not Interruptible**. The person is very busy, the robot should not interrupt.

INT-0 **Interruptibility Unknown**. The robot is aware that a person is present, but does not have sufficient sensory input to analyze interruptibility.

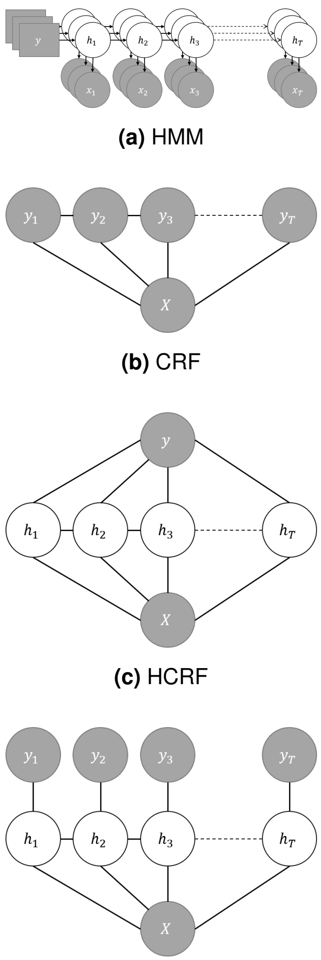
Values 1–4 in the scale capture the full range of interruptibility states that can help guide the robot's decision making process. We include the rating of 0 to represent states in which the robot does not yet have sufficient information about the person. In this case, the robot may choose to approach another person or take actions to improve its sensing quality.

Features for Interruptibility

Combining prior work in robotics [2, 8] with Ubicomp [17], we estimate interruptibility through sensory features that describe the *person state*, and the *interruption context*.

Person state is inferred from laser, video, and audio sensor data and represents three information categories. The first category is the position and orientation of a person's body in the environment with respect to the robot. The second category is the head orientation and the gaze direction of the person. The final category includes information on the presence and orientation of sound in the environment. In light of noisy sensor data, we created three feature sets consisting of subsets of features from these categories—*Minimal Features Set* (Min), *Standard Features Set* (Std), and *Extended Features Set* (Ext)—to explore the tradeoff between noise and information on classification performance for interruptibility. From Min to Ext, the number of features used for classification increased concomitantly with the amount of noise in the features.

Interruption context can be inferred from many factors, including known information about the user, the task, the environment and the type of interruption [17]. In this work, we consider environmental (or scene) context, which we define



(a) HMM

(b) CRF

(c) HCRF

(d) LDCRF

Figure 2: Graphical representation of each of the computational models in this paper. Gray elements represent observed variables, and white elements represent hidden variables

as visually observable cues that may inform the robot about the interruptibility of a person. In particular, we propose that objects the person is interacting with can serve as useful visual context cues. For example, an individual nursing a coffee mug in a lounge is judged to be more interruptible than someone engaged with a laptop in the same lounge. Although objects cannot capture all of the complexities of interruption context, object recognition is widely available on robotic systems and we hypothesize that, combined with traditionally used cues of person state, object labels can improve the estimate of a person's interruptibility. Thus we add the object labels to our perception model.

Models for Interruptibility

Following the example of prior work, we used temporal models to estimate interruptibility. Mollaret et al. [8] and Chiang et al. [2] both used Hidden Markov Models (HMMs) to address related problems, with promising results, and so we adopt this model as our baseline. Additionally, we explore the use of Conditional Random Fields (CRFs) [6] and derivatives thereof, Hidden Conditional Random Fields (HCRFs) [16], and Latent-Dynamic Conditional Random Fields (LDCRFs) [9], as alternate temporal models to classify interruptibility. We hypothesize that CRFs will outperform HMMs in classifying interruptibility due to their more expressive representation using feature functions that can be dependent on the entire observation sequence. We also hypothesize that HCRFs and LDCRFs will perform better than the CRFs due to their ability to model intra-class variation within observed data through hidden states.

Data Collection and Annotation

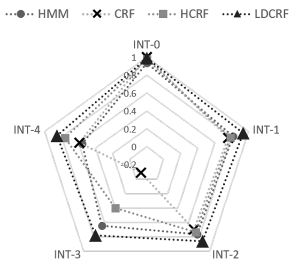
To evaluate the performance of the above models on classifying interruptibility, we performed data collection to obtain videos of small groups of people in a public space. We asked five people (not co-authors on the paper) to enact everyday activities in a common area of the building. The robot was teleoperated through five data collection runs through a preset series of waypoints that enabled it to observe the group from different perspectives; each run lasted an average of 108 seconds. The data from each of the runs was segmented into 250 ms non-overlapping time windows in order to perform rudimentary sensor fusion using Euclidean distance heuristics.

We annotated each of the data segments with interruptibility estimates and verified our labels by calculating a Cronbach's Alpha measure of inter-rater reliability with two independent coders. These coders were asked to annotate a random subset of approximately 40% of the data, and the resulting scores of 0.81 and 0.96 indicated a high level of agreement. In addition, the segments were annotated with object labels from within the set *unknown, none, laptop, bottle, book, headphones, mug, phone_talk,* and *phone_text* to simulate automated object recognition. We swapped the object labels for 10% of the segments in each of the interruptibility classes to simulate noise in object recognition.

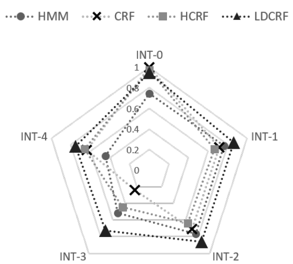
The annotated data segments were finally concatenated into sequences of minimum length 4 (1 second) and maximum length 8 (2 seconds) in preparation for model evaluation. In the event of missing data in a sequence, values were imputed either through linear interpolation (continuous) or by propagating the last known value (boolean). If neither approach was available, attributes were assigned a value of NaN to distinguish them from other valid values in the domain.

Results

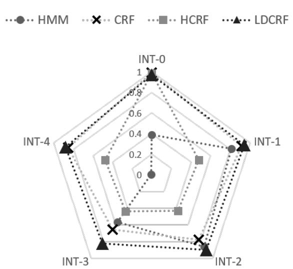
We evaluated our four candidate models using 10 fold cross-validation with 80% of the data in a fold used for training and 20% for testing. Results with the best performing pa-



(a) Minimal Feature Set (Min)



(b) Standard Feature Set (Std)



(c) Extended Feature Set (Ext)

Figure 3: Radar plots reporting MCC performance of each model as a function of the interruptibility class.

rameters for each model are reported here using Matthew's Correlation Coefficient (MCC).

Model Analysis
Fig. 3 compares the performance of the HMM, CRF, HCRF and LDCRF models across the three feature sets without the inclusion of object context data. We find two major trends in the data. Firstly, adding more information through more features has the potential to improve the decision making ability of the models with the LDCRF and CRF showing the most improvement. In fact, the MCC score of the CRF in classifying INT-3 improves from -0.08 (slightly worse than random), with *Min* features, to 0.65, with *Ext* features. Secondly, we find that the LDCRF and CRF prove robust to the noise in the added features while the classification performance of the HCRF and HMM suffer greatly. For example, the MCC score for the HMM in classifying INT-4 drops from 0.55, with *Min* features, to 0.01 (no better than random), with *Ext* features.

All in all, we find that the LDCRF model consistently outperforms all other methods across all feature sets, with the best performance achieved on the *Ext* features. This result indicates the robustness of the LDCRF to noisy data, which is valuable given the expected variability in the quality of data available to a mobile robot in public spaces.

Object Context
In this section, we evaluate the effect of adding object recognition features to each of the three feature sets on classification performance. Given the dominant performance of LDCRFs in the previous section, we report analysis of only the LDCRF model on these datasets, although, we observe similar effects across the other three temporal models.

Fig. 4 presents a comparison of LDCRF performance on the original feature sets (black) and with the addition of object labels (gray). As can be seen, the addition of object labels consistently increases the classification performance of the model across nearly all conditions. The only exceptions are for *Ext+Obj* in INT-4 and *Min+Obj* in INT-0, in which there is a negligible loss of 0.001 and 0.005 in MCC scores respectively. In all other conditions we observe an increase in performance, particularly for INT-3 where the MCC score improves by as much as 0.16 points.

The best overall performance is achieved by *Min+Obj* condition supporting our hypothesis that contextual information derived from object labels is highly informative to interruptibility classification. Furthermore, we observe a tradeoff between the the use of a larger set of, somewhat noisy, features (*Ext*) and the use of a smaller number of more precise features. Specifically, in the absence of object labels, LDCRF performance is highest with the *Ext* features, making the best use of the additional information, even when it is noisy. With the introduction of object labels, the *Ext* features serve as a distraction and the best performance is achieved with the *Min* features. This finding is significant in guiding future development efforts in this area. Specifically, we observe that all domains, but especially ones in which it is relatively difficult to obtain reliable person tracking information, benefit from the incorporation of contextual signals.

Conclusion
In this paper, we introduced a rating scale for characterizing interruptibility, and compared four temporal models – HMMs, CRFs, HCRFs and LDCRFs – in classifying the interruptibility of multiple people in a scene based on laser, visual and audio data collected by a mobile robot. Our findings show that LDCRFs consistently outperform other models across all conditions. Additionally, our work is the first to introduce contextual scene information beyond the person-of-interest, in this case object labels, to models of interrupt-

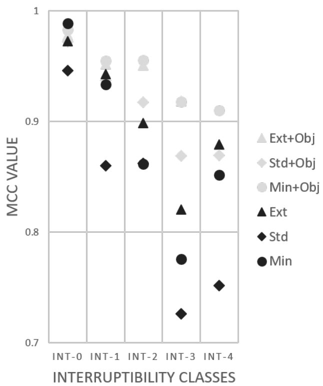


Figure 4: Effect of adding object labels to LDCRF.

ibility. Our findings show that adding object labels significantly improves interruptibility classification performance, particularly when combined with reliable person descriptive features. Our approach successfully handles multiple people in a single scene, and in future work we will explore how the presented interruptibility ratings can be used by the robot to decide who to interrupt, and how.

Acknowledgments

This work was supported by an Early Career Faculty grant from NASA's Space Technology Research Grants Program.

REFERENCES

1. Piotr D. Adamczyk and Brian P. Bailey. 2004. If not now, when?: the effects of interruption at different moments within task execution. In *Proc. SIGCHI conf. on Human factors in comp. sys.* ACM Press, 271–278.
2. Yi-Shiu Chiang, Ting-Sheng Chu, Chung Dial Lim, Tung-Yen Wu, Shih-Huan Tseng, and Li-Chen Fu. 2014. Personalizing robot behavior for interruption in social human-robot interaction. In *Int. Work. on ARSO*. IEEE, 44–49.
3. James Fogarty, Scott E. Hudson, Christopher G. Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny C. Lee, and Jie Yang. 2005. Predicting human interruptibility with sensors. *Tran. on Comp. Human Int.* 12, 1 (2005), 119–146.
4. Tony Gillie and Donald Broadbent. 1989. What makes interruptions disruptive? A study of length, similarity, and complexity. *Psych. Research* 50, 4 (1989), 243–250.
5. Eric Horvitz and Johnson Apacible. 2003. Learning and reasoning about interruption. In *Proc. Int. Conf. on Multimodal Interfaces*. ACM, 20–27.
6. John Lafferty, Andrew McCallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. Int. Conf. on Machine Learning*, Vol. 1. 282–289.
7. Daniel McFarlane and Kara Latorella. 2002. The Scope and Importance of Human Interruption in Human-Computer Interaction Design. *Human-Computer Interaction* 17, 1 (2002), 1–61.
8. C. Mollaret, A.A. Mekonnen, F. Lerasle, I. Ferrané, J. Pinquier, B. Boudet, and P. Rumeau. 2016. A multi-modal perception based assistive robotic system for the elderly. *Computer Vision and Image Understanding* (2016).
9. Louis-Philippe Morency, Ariadna Quattoni, and Trevor Darrell. 2007. Latent-Dynamic Discriminative Models for Continuous Gesture Recognition. In *Conf. on CVPR*. IEEE, 1–8.
10. A. Joy Rivera. 2014. A socio-technical systems approach to studying interruptions: Understanding the interrupter's perspective. *Applied Ergonomics* 45, 3 (2014), 747–756.
11. Stephanie Rosenthal, Manuela M. Veloso, and Anind K. Dey. 2012. Is Someone in this Office Available to Help Me? *Journal of Intelligent & Robotic Systems* 66, 1-2 (2012), 205–221.
12. Satoru Satake, Takayuki Kanda, Dylan F. Glas, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. 2009. How to approach humans?-strategies for social robots to initiate interaction. In *Int. Conf. on HRI*. IEEE, 109–116.
13. Paul Saulnier, Ehud Sharlin, and Saul Greenberg. 2011. Exploring minimal nonverbal interruption in HRI. In *RO-MAN*. IEEE, 79–86.
14. Chao Shi, Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2015. Measuring Communication Participation to Initiate Conversation in Human-Robot Interaction. *International Journal of Social Robotics* 7, 5 (2015), 889–910.
15. Hermann Stern, Viktoria Pammer, and Stefanie N Lindstaedt. 2011. A preliminary study on interruptibility detection based on location and calendar information. *Proc. CoSDEO* 11 (2011).
16. Sy Bor Wang, A. Quattoni, L.-P. Morency, D. Demirdjian, and T. Darrell. 2006. Hidden Conditional Random Fields for Gesture Recognition. In *Conf. on CVPR*, Vol. 2. IEEE, 1521–1527.
17. Liam D. Turner, Stuart M. Allen, and Roger M. Whitaker. 2015. Interruptibility prediction for ubiquitous systems. In *Proc. Int. Conf. on Pervasive and Ubiquitous Comp.* ACM, 801–812.